DOCUMENT RESUME

ED 274 681 TM 860 541

AUTHOR Vale, C. David; Gialluca, Kathleen A.
TITLE ASCAL: A Microcomputer Program for Estimating Logistic IRT Item Parameters.
INSTITUTION Assessment Systems Corp., St. Paul, Minn.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
REPORT NO ONR-RR-85-4
PUB DATE 11 Nov 85
CONTRACT 00014-83-C-0634
NOTE 31p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adaptive Testing; Bayesian Statistics; *Computer Assisted Testing; Computer Simulation; Correlation; *Estimation (Mathematics); Evaluation Criteria; *Item Analysis; Item Banks; *Latent Trait Theory; Microcomputers; *Monte Carlo Methods; Programing Languages; Test Theory
IDENTIFIERS Armed Services Vocational Aptitude Battery; *ASCAL Computer Program; *Calibration; Item Parameters; Logist Computer Program

ABSTRACT

ASCAL is a microcomputer-based program for calibrating items according to the three-parameter logistic model of item response theory. It uses a modified multivariate Newton-Raphson procedure for estimating item parameters. This study evaluated this procedure using Monte Carlo Simulation Techniques. The current version of ASCAL was then compared to the current version of LOGIST and a prior version of ASCAL that was used to estimate the parameters in the adaptive item pool of the Armed Services Vocational Aptitude Battery. Parameters were estimated for the items in each of three different-simulated tests. Three different evaluative criteria were used: (1) root mean squared error between true and estimated item parameters, (2) the correlation between true and estimated item parameters, and (3) calibration efficiency. Results suggest that ASCAL produces parameter estimates that are at least as accurate as those produced by LOGIST. There were only minor differences in the three evaluative criteria between the two versions of ASCAL. The differences between ASCAL and LOGIST were slightly larger than the differences between the two versions of ASCAL. Item parameter error increased markedly according to all three criteria and for all calibration procedures when items with an exaggerated range of difficulty parameters were calibrated. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

unclassified
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT approved for public release; distribution unlimited
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) ONR-85-4	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Assessment Systems Corporation 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) 2233 University Avenue, Suite 310 St. Paul, MN 55114	7b. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street, Code 442 Arlington, VA 22217
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Personnel and Training Research 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-83-C-0634
8c. ADDRESS (City, State, and ZIP Code) Office of Naval Research 800 N. Quincy Street, Code 442 Arlington, VA 22217	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. N 507-002 / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
ASCAL: A Microcomputer Program for Estimating Logistic IRT Item Parameters

12. PERSONAL AUTHOR(S)
C. David Vale & Kathleen A. Gialluca

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
technical report	FROM 9/1/83 TO 1/11/85	85 Nov 11	17

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	adaptive testing, ASCAL, item analysis, item banking, item calibration, IRT, item response theory, LOGIST, MicroCAT, parameter estimation, tailored testing, test analysis

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

ASCAL is a microcomputer-based program for calibrating items according to the three-parameter logistic model of item response theory. ASCAL employs Lord's (1974) modified likelihood equations (for items that are omitted or not reached) and Bayesian prior distributions on the discrimination, guessing, and ability parameters to arrive at final estimates of the item parameters. No ability parameters are produced.

ASCAL uses a modified multivariate Newton-Raphson procedure for estimating item parameters. The estimation process begins by specifying starting points for the ability and item parameters. In this procedure, abilities are first estimated using the preliminary estimates of the item parameters. After ability estimates have been obtained for all examinees, they are sorted and grouped into

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION	
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Charles E. Davis	22b. TELEPHONE (Include Area Code) 202-696-4046	22c. OFFICE SYMBOL

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
unclassified

Block 16. (Continued)

20 fractiles with approximately equal numbers of examinees in each. The item parameters are then estimated by assuming that the 20 fractile means are representative of the entire ability distribution. The sequence of ability estimation, ability grouping, and parameter estimation is repeated until the item parameters converge on stable values or fail to improve.

This procedure was evaluated using Monte Carlo simulation techniques. The current version of ASCAL was then compared to the current version of LOGIST (Wingersky, Barton, & Lord, 1982) and a prior version of ASCAL that was used to estimate the parameters in the adaptive item pool of the Armed Services Vocational Aptitude Battery. Parameters were estimated for the items in each of three different simulated tests. Three different evaluative criteria were used: (1) root mean squared error between true and estimated item parameters, (2) the correlation between true and estimated item parameters, and (3) calibration efficiency. This last criterion measures the amount of information in the estimated parameters relative to the information in the true parameters and is an indication of the joint effects of calibration error in the individual parameters.

The results of this evaluation suggest that ASCAL produces parameter estimates that are at least as accurate as those produced by LOGIST. There were only minor (and inconsequential) differences in the three evaluative criteria between the two versions of ASCAL. The differences between ASCAL and LOGIST were slightly larger than the differences between the two versions of ASCAL; however, even these differences were small. Item parameter error increased markedly (for each of the three parameters) according to all three criteria and for all calibration procedures when items with an exaggerated range of difficulty parameters were calibrated. Only the difficulty parameter correlations were unaffected by this manipulation.

ABSTRACT

ASCAL is a microcomputer-based program for calibrating items according to the three-parameter logistic model of item response theory. ASCAL employs Lord's (1974) modified likelihood equations (for items that are omitted or not reached) and Bayesian prior distributions on the discrimination, guessing, and ability parameters to arrive at final estimates of the item parameters. No ability parameters are produced.

ASCAL uses a modified multivariate Newton-Raphson procedure for estimating item parameters. The estimation process begins by specifying starting points for the ability and item parameters. In this procedure, abilities are first estimated using the preliminary estimates of the item parameters. After ability estimates have been obtained for all examinees, they are sorted and grouped into 20 fractiles with approximately equal numbers of examinees in each. The item parameters are then estimated by assuming that the 20 fractile means are representative of the entire ability distribution. The sequence of ability estimation, ability grouping, and parameter estimation is repeated until the item parameters converge on stable values or fail to improve.

This procedure was evaluated using Monte Carlo simulation techniques. The current version of ASCAL was then compared to the current version of LOGIST (Wingersky, Barton, & Lord, 1982) and a prior version of ASCAL that was used to estimate the parameters in the adaptive item pool of the Armed Services Vocational Aptitude Battery. Parameters were estimated for the items in each of three different simulated tests. Three different evaluative criteria were used: (1) root mean squared error between true and estimated item parameters, (2) the correlation between true and estimated item parameters, and (3) calibration efficiency. This last criterion measures the amount of information in the estimated parameters relative to the information in the true parameters and is an indication of the joint effects of calibration error in the individual parameters.

The results of this evaluation suggest that ASCAL produces parameter estimates that are at least as accurate as those produced by LOGIST. There were only minor (and inconsequential) differences in the three evaluative criteria between the two versions of ASCAL. The differences between ASCAL and LOGIST were slightly larger than the differences between the two versions of ASCAL; however, even these differences were small. Item parameter error increased markedly (for each of the three parameters) according to all three criteria and for all calibration procedures when items with an exaggerated range of difficulty parameters were calibrated. Only the difficulty parameter correlations were unaffected by this manipulation.

ACKNOWLEDGEMENTS

We would like to thank J. Stephen Prestwood and James B. Sympson for their helpful suggestions in the development of ASCAL and Fred Lord, Malcolm Ree, Martha Stocking, and Marilyn Wingersky for their assistance in the comparisons with LOGIST.

TABLE OF CONTENTS

Introduction 1

Three-Parameter Logistic IRT Model 1

Bayesian Estimation of Ability and Item Parameters 2

Numerical Implementation 4

Overview 4

Initial Estimates 5

Ability Estimation 5

Ability Grouping 5

Parameter Estimation 6

Program Termination 7

Program Operation 7

Input 7

Output 8

System requirements 8

Program Evaluation 8

Method 8

Simulated item responses 8

Calibration programs 9

Evaluative criteria 9

Theta-gamma transformation 10

Efficiency 12

Results 13

RMSE 13

Correlations 14

Efficiency 15

Summary 16

Discussion 16

References 17

INTRODUCTION

The MicroCAT^{tm} Testing System is a self-contained system of programs for developing, administering, and evaluating psychological tests using both adaptive and conventional strategies (Assessment Systems Corporation, 1984). The system runs on IBM personal computers, either individually or in a local area network. MicroCAT^{tm} uses item response theory (IRT; Lord & Novick, 1968) to calibrate adaptive test items. The program within the MicroCAT^{tm} system that performs IRT item calibration is called ASCAL.

ASCAL was the first IRT calibration program for microcomputers. The most popular IRT calibration program, LOGIST (Wingersky, Barton, & Lord, 1982), was available only for mainframe-class machines and was typically quite expensive to run. It was not unusual for the cost of calibrating the items in a single test to exceed $50 for computer time alone.

Three-Parameter Logistic IRT Model

Both LOGIST and ASCAL estimate parameters for the three-parameter logistic IRT model, which was developed for use with dichotomously scored multiple-choice items. The item response function in this model expresses the probability of a correct response as a function of examinee ability, theta (O), and three item parameters, a, b, and c. The item response function is an S-shaped curve going from a lower asymptote equal to the c parameter (the pseudo-guessing parameter) to a maximum value of 1.0. The midpoint of this curve has its projection onto the ability scale at b (the item difficulty parameter). The slope or rate at which the probability increases as a function of ability is a function of the a parameter (the discrimination parameter). The three-parameter IRT model is specified in Equation 1.

$$P_g(\theta_i) = \text{Prob}\,(U_{gi} = 1 \mid a_g, b_g, c_g, \theta_i) \qquad [1]$$
$$= c_g + (1 - c_g)\ \Psi[Da_g(\theta_i - b_g)]$$

where

the subscripts i and g index the examinee and item, respectively;

$$U_{gi} = \begin{cases} 1 \text{ if examinee } i \text{ answered item } g \text{ correctly} \\ 0 \text{ otherwise;} \end{cases}$$

$\Psi(x) = 1 / (1 + \exp(-x))$; and

$D = 1.7$.

LOGIST's method for estimating item parameters is based on an adaptation, proposed by Lord (1974), of the maximum likelihood principle. This model makes allowances for items that are unanswered or omitted by an examinee. When all examinees respond to all items in the test, Lord's procedure is equivalent to the classical maximum likelihood procedure.

Lord made a distinction between items that were *omitted* and those that were *not reached*. Unanswered items embedded within the sequence of items to which the examinee did respond were considered omitted items. That is, it was assumed that the examinee had the opportunity to respond to the items but chose, for some reason, to skip or omit those items. On the other hand, unanswered items at the end of a test were considered not reached. In this case, it was assumed that the examinee did not have enough time to respond to all of the items in the test. Similarly, items that are never presented to some subset of examinees can be considered not reached for those examinees.

In Lord's scheme, unreached items are not used to estimate an examinee's ability; the scored response vector does not include the block of unanswered items that occurs at the end of a test. To score unreached items as incorrect would mean that an examinee's score is dependent, at least partially, on the speed with which he or she answered the items on the test. This is a major violation of the assumptions of IRT, which models the probability of a correct answer on item characteristics and the examinee's ability only.

On the other hand, examinees are given partial credit for those items that they chose to omit. The rationale for this is twofold: (1) it is assumed that the examinees were administered these items and had ample time to respond to them, and (2) they would have answered some of these items correctly if they had guessed rather than skipped the item.

Bayesian Estimation of Ability and Item Parameters

Maximum likelihood procedures for estimating item parameters have several desirable characteristics (for example, under certain conditions, they produce estimates that are asymptotically efficient and asymptotically normally distributed). In practice, however, such procedures are often difficult to implement. For example, no maximum likelihood ability estimate can be obtained for an examinee if he or she answers all of the items correctly or answers fewer items correctly than would be expected by chance. Furthermore, for complicated estimation situations (such as item calibration), numerical compromises must often be made; a number of ad hoc decision rules and auxiliary estimation procedures were incorporated into LOGIST during its development. For example, estimates of the c parameter cannot be obtained by LOGIST for items that are too easy. This is because typical data sets do not provide sufficient information to allow such parameters to be estimated by the maximum likelihood method with any accuracy. Thus, LOGIST will set the c parameters of such items to the average c parameter of the remaining items.

The calibration procedures of ASCAL were modeled after those in LOGIST. However, ASCAL makes a different set of numerical compromises that allow, in effect, a conceptually simpler approach to item calibration. That is, ASCAL

adds Bayesian prior distributions to the pseudo-likelihood functions for the ability estimates and the *a* and *c* parameters (no prior distribution is used for the *b* parameter). Thus, the pseudo-maximum likelihood estimation process of LOGIST is a pseudo-Bayesian estimation process in ASCAL. The prior distribution of ability chosen for ASCAL is a standard normal distribution. By convention, all three-parameter IRT calibration programs set the scale of ability to have a mean of zero and a variance of one; the normal shape of the prior distribution used here was chosen as the most representative general form for a distribution of ability.

Symmetric beta distributions were used as the specified Bayesian priors for the *a* and *c* parameters. These distributions were chosen because they provided a continuous bounding mechanism consistent with intuition concerning probable values of the *a* and *c* parameters. The prior distribution for the *a* parameters was specified as $B(3.0,3.0)$ with upper and lower bounds of 2.60 and 0.30, respectively. The prior distribution for the *c* parameters was specified as $B(5.0,5.0)$ with a lower bound of -0.05 and an upper bound equal to 0.05 plus twice the reciprocal of the number of alternatives. The classical beta density functions were redefined (using a simple change-of-variable transformation) to have the bounds specified above instead of upper and lower bounds of one and zero, respectively (see below). These transformations ensured that the *a* and *c* parameters were in the appropriate range. The estimates of the *b* parameters were bounded by $\pm$ 3.00.

The criterion function maximized by this pseudo-Bayesian method is shown in Equation 2. This is the pseudo-likelihood function proposed by Lord (1974; Wood, Wingersky, & Lord, 1976, Equation 4) weighted by a univariate normal distribution on theta and univariate beta distributions on *a* and *c*.

$$L^* = \prod_{i=1}^{N} \prod_{g=1}^{n_i} P_g(\theta_i)^{v_{gi}} Q_g(\theta_i)^{1-v_{gi}} \varphi(\theta_i) * \qquad [2]$$

$$f(a_g, 3.0, 3.0, 0.3, 2.6) *$$

$$f(c_g, 5.0, 5.0, -0.05, (0.05 + 2/K))$$

where

n_i is the number of items that were answered by examinee i;

$Q_g(\theta_i) = 1 - P_g(\theta_i)$;

$$v_{gi} = \begin{cases} u_{gi} & \text{if examinee } i \text{ responded to item } g \\ 1/K & \text{otherwise, where } K \text{ is the number of alternatives;} \end{cases}$$

φ is the standard normal density function;

$$f(x,r,s,j,k) = \frac{1}{B(r,s)} * \frac{1}{(k - j)} * \left[\frac{x - j}{k - j}\right]^{r-1} * \left[1 - \frac{x - j}{k - j}\right]^{s-1}$$

for upper and lower bounds k and j, respectively, and beta-function parameters r and s;

$$\text{and } B(r,s) = \int_0^1 y^{r-1} (1 - y)^{s-1} \, dy$$

is the classical form of the beta function.

Simultaneous solution for the roots of the derivative equations for all parameters is numerically intractable. Therefore, ASCAL uses the general iterative procedure followed by LOGIST. In this procedure, the ability parameters and the item parameters are estimated separately. First, the abilities are estimated while the item parameters are assumed to be known; then the item parameters are estimated while the abilities are assumed to be known. This process is repeated, and the ability and item parameter estimates are updated and refined at each stage. While there is no guarantee that this process will converge on final ability and item parameter estimates, the experience with LOGIST and ASCAL has been that acceptable estimates are produced.

NUMERICAL IMPLEMENTATION

Overview

ASCAL uses a modified multivariate Newton-Raphson procedure for estimating the parameters. The estimation process begins by specifying starting points for the ability and item parameters. Once these starting points have been obtained, the modified Newton-Raphson procedure begins. In this modified procedure, abilities are first estimated using the preliminary estimates of the item parameters.

After ability estimates have been obtained for all examinees, they are sorted and grouped into 20 fractiles with approximately equal numbers of examinees in each; this grouping is done for computational convenience. After grouping, the fractile means are weighted by the number of subjects in each fractile and then standardized. The mean ability for each fractile is taken as representative of all examinees contained in that fractile. The item parameters

are then estimated by assuming the 20 ability levels to be representative of the entire ability distribution.

The sequence of ability estimation, ability grouping, and parameter estimation is repeated several times until the item parameters converge on stable values or fail to improve. The details of each of these processes are described below.

Initial Estimates

A Newton-Raphson iteration requires that initial estimates be specified for all of the parameters that are to be estimated. These estimates must be reasonably accurate or the iteration process may diverge and fail to produce acceptable estimates.

The initial ability estimates are obtained from raw formula scores. These scores are computed as the number of items answered correctly minus a fraction of the items answered incorrectly. For each incorrect item, the fraction subtracted is equal to the reciprocal of one less than the number of alternatives. These formula scores are then standardized using a linear transformation based on the mean and standard deviation of the formula scores in the group of examinees.

The initial item parameters are obtained through heuristic transformations of the classical item statistics as described by Jensema (1976). The *c* parameters are estimated as the reciprocal of the number of alternatives for each item. The initial estimates of the *a* and *b* parameters are then computed from the corrected biserial item-total correlation and the proportion of examinees answering the item correctly.

In the first of the iterations that follow, the initial item-parameter estimates are used to obtain the first estimates of ability. The initial ability parameters obtained by standardizing the formula scores are used only as starting points for the Newton-Raphson iteration.

Ability Estimation

Ability estimates are obtained through Newton-Raphson iteration on the derivative with respect to theta of the criterion function L^* shown in Equation 2. Using both the first and second derivatives of the criterion function, the iteration proceeds for each examinee individually until an ability value is found for that examinee that results in a root for the first derivative of the criterion function. The estimate thus obtained is very similar to a Bayesian modal estimator of ability. It differs from a true Bayesian modal estimator in that this estimator incorporates Lord's pseudo-likelihood function rather than a true likelihood function.

Ability Grouping

When the item parameters are estimated, the criterion function must be summed over all ability levels. This could be done for each examinee

individually or the examinees could be grouped into fractiles and the sum could then be taken over these groups. The latter procedure results in considerable computational savings, and is the procedure employed by ASCAL, using 20 fractiles. There is precedent for grouping abilities in this way; such a procedure is used both in LOGIST and in a related calibration program, LOGOG (Kolakowski and Bock, 1973). LOGIST documentation recommends using grouped abilities only in some stages of estimation. Like LOGOG, ASCAL uses grouped abilities for all stages of estimation.

The ability-grouping process begins by sorting abilities from low to high. ASCAL then attempts to group these abilities into 20 fractiles, each containing an equal number of examinees. To accomplish this, the ability level for the examinee at each 20th of the distribution is obtained as a bound. Examinees are then grouped into the fractiles with a bound most immediately above their ability levels. This process results in equal-sized groups except when examinees with equal abilities span a fractile boundary or when the number of examinees is not evenly divisible by 20. Typically, neither of these conditions results in any substantial differences in fractile sizes. Fractile divisions are done on a running total of sample fractions so that round-off error does not result in an excessive number of subjects in the last fractile.

The mean ability is then computed in each fractile and taken as the representative ability for all examinees in that fractile. Whereas the median ability in the fractile might be more appropriate for a procedure such as LOGOG or LOGIST which uses maximum likelihood estimates of ability (with the potential of infinite ability estimates on finite length tests) the mean ability is an appropriate characterization for ASCAL in which the abilities are obtained by a Bayesian method.

The 20 fractile means represent all ability levels in the group of examinees. These 20 ability levels are then standardized so that they have a mean of zero and a standard deviation of one (as was assumed for the ability distribution). ASCAL performs the standardization by weighting each fractile mean by the number of examinees in the fractile. If the grouping procedure produced exactly equal numbers of examinees in each of the fractiles, this weighting would be unnecessary. The weighting is done to preclude complications in cases where the numbers of examinees are not exactly equal.

Parameter Estimation

The item parameters are estimated for one item at a time. The *a* and *b* parameters are estimated using a Newton-Raphson procedure. The *c* parameters are estimated by systematically stepping through possible values. Thus the *a* and *b* estimation procedure is performed within a loop that steps through all reasonable values of the *c* parameter.

For computational efficiency, the *c* parameter stepping process is done in two stages. In the first stage, trial values of *c* are evaluated starting at the *c* estimate from the previous loop and stepping outward in steps of 0.05. The process first steps down until a limit is reached or the criterion function decreases. Then, if no increase was observed while stepping down, the stepping

process steps up until the criterion function decreases or a limit is reached. A second stage of iteration is then performed starting at the best *c* value obtained in the first stage and proceeding outward (in steps of 0.01) to a maximum distance of 0.04 from the second-stage starting point or until the criterion function fails to increase.

At each trial value of *c*, the *a* and *b* parameters are estimated. These parameters are estimated jointly using a bivariate Newton-Raphson procedure if the matrix of partial second derivatives is positive definite. Otherwise, the *a* and the *b* parameters are estimated independently. To prevent wild fluctuations in the *a* and *b* parameters during the estimation process, the maximum adjustment to the *a* parameters is limited to plus or minus 0.2 and the maximum adjustment to the *b* parameters is limited to plus or minus 0.4. The *a* and *b* parameters are considered to have converged for a trial value of *c* if the sum of the absolute change in *a* and the absolute change in *b* is less than 0.01.

If the *a* and *b* parameters have failed to converge after 10 iterations, the change indicated for each parameter by the Newton-Raphson procedure is reduced by a power of 0.8 for each additional iteration. Thus, the prescribed change is dampened or multiplied by 0.8 on the eleventh iteration, by 0.64 on the twelfth iteration, etc.

Program Termination

ASCAL terminates its iterative process either when the parameter estimates converge on constant values or when the maximum number of iterations allowed by the user is reached. The item parameters are assumed to have converged if the sum of the absolute changes in *a*, *b*, and *c* from one iteration to the next is less than 0.01 for two consecutive iterations on every item. Alternatively, if the total number of loops reaches the maximum number allowed by the user, the program terminates.

Program Operation

The details of program operation are described in Chapter 12 of the *User's Manual for the MicroCAT Testing System* (Assessment Systems Corporation, 1984). ASCAL was designed to estimate item parameters only and was not intended to be a scoring program; thus, no ability estimates are provided for the examinees. Because no editing capabilities are included in the program, it is assumed that all data are input to ASCAL in correct and final form.

Input. The input data file contains raw item responses for each examinee. This data file must also contain the item-specific information required for item calibration; this additional information is included in the data file and not in a separate program-control file. In addition to the raw item responses, then, the input data file contains the following information: (1) number of items, (2) symbol codes for omitted and unreached items, (3) rudimentary formatting information indicating how many columns of the examinee response record contain identification data, (4) the number of response alternatives for each item, (5) the keyed response for each item, and (6) a flag for each item indicating whether or not it should be included in the analysis.

The interactive operation of ASCAL is quite simple. The user needs only to provide the name of the input file, the name of the output file, and the maximum number of loops through which the program should proceed, if it has not converged prior to that point. Ten iterations usually provide adequate estimates.

Output. ASCAL output is a simple listing of the following information: (1) the user-specified program-control parameters, (2) the heuristic item parameters used as initial starting values, (3) parameter changes at each stage through the parameter-estimation process, and (4) the final parameters and their Pearson chi-square lack-of-fit test results.

System requirements. ASCAL runs on an IBM PC, PC XT, or PC AT and many of the IBM PC-compatible computers. Memory requirements for ASCAL are somewhere between 128K and 192K. ASCAL will run on systems either with or without the math coprocessor chip; however, the run time without this chip can be excessive. For example, for a 50-item test administered to 2,000 examinees, the typical time with the coprocessor chip is approximately 2 hours; without the coprocessor chip, the same run may take more than 24 hours. In either case, the results will be identical.

PROGRAM EVALUATION

Method

ASCAL was evaluated using a Monte Carlo simulation (see Ree, 1978, or Vale & Weiss, 1975, for a full description of a simulation). In this study, scored responses to items with known ("true") *a*, *b*, and *c* parameters were generated using the three-parameter logistic IRT model. ASCAL was then used to estimate item parameters from these simulated item responses. The estimated item parameters were compared with the true item parameters using several different criteria. This simulation procedure is discussed in more detail below.

Simulated item responses. Three sets of true parameters were used. The first set of parameters was obtained from a 25-item test containing general science items. To obtain an effective test length of 50, each item and its parameters were included in the test twice. All items in Test 1 had four alternatives. Test 1 had a restricted range of item difficulty, typical of a conventional test, but inappropriate for adaptive testing. To create a test more appropriate for adaptive testing, Test 2 was created by multiplying all of the *b* parameters in Test 1 by 2.0. Thus Tests 1 and 2 were identical except for the *b* parameters. Test 3 was modeled after a 57-item test of shop knowledge. Each item in Test 3 had five alternatives. Since Test 3 was developed as part of an adaptive item pool, it had a slightly wider range of difficulty than did Test 1, although its difficulty range was not as wide as that of Test 2. Table 1 presents the means and standard deviations of the true parameters used as models for the simulations.

Item response data were generated for 2,000 examinees for each of the three tests. Examinee ability levels were sampled from a standard normal distribution.

Calibration programs. Although the objective of parameter estimation is to obtain estimates that are identical to the true values, in practice this never happens. Thus, to provide a basis for comparison, the current version of LOGIST (LOGIST5; Wingersky, et al., 1982) was also used to estimate parameters from these data.

Table 1
Means and Standard Deviations of the True Item Parameters for Tests 1, 2, and 3

	Test 1		Test 2		Test 3	
Parameter	Mean	S.D.	Mean	S.D.	Mean	S.D.
a	1.492	0.410	1.492	0.410	1.041	0.417
b	-0.090	0.910	-0.179	1.820	0.675	1.069
c	0.230	0.082	0.230	0.082	0.198	0.028

Two versions of ASCAL were used to estimate parameters on these data. Version 2.0 is the version of ASCAL that was used for estimation of the parameters of the item pools for the computerized adaptive version of the Armed Services Vocational Aptitude Battery (ASVAB). Since that calibration, several numerical improvements were made to ASCAL resulting in the current version, Version 3.0. Both versions were used here to investigate the magnitude of the differences between the two versions.

Evaluative criteria. Three different criteria were used to evaluate the accuracy of the item calibration programs. The root mean squared error (RMSE, the square root of the average squared difference between the true and estimated parameters) was computed for each of the three (a, b, and c) parameters in each of the three tests. Similarly, the Pearson product-moment correlation coefficient between the estimated and the true parameters was computed for each parameter in each test. The third criterion was a calibration efficiency criterion.

The efficiency criterion is a relative-information criterion suggested by Vale, Maurelli, Gialluca, Weiss, and Ree (1981). Their procedure computes the amounts of psychometric information (Birnbaum, 1968) that would be extracted from the items if they were scored using the estimated or errant parameters; the relative efficiency of the estimated parameters (i.e., the ratio of the information in the estimated parameters to the information in the true parameters) can be determined for each theta level by comparing the errant information with the true information. This process is described below.

Theta-gamma transformation. The theta metric can be defined as the criterion metric along which the true parameters are anchored and along which the response probabilities are accurately described by the IRT model incorporating the theta level and the item parameters. A second metric, call it gamma, is produced by scoring item responses using parameters other than the true parameters of the theta metric (i.e., by using the errant or estimated parameters). The gamma metric, then, is a one-to-one transformation of the theta metric. The gamma level corresponding to a given theta level could be determined, conceptually, from administering a test scored using the errant parameters an infinite number of times and scoring all the responses simultaneously. Each theta value would then asymptotically converge on a single gamma value.

It is, of course, impossible to administer an infinite-length test or to repeat a finite-length test an infinite number of times. However, the theta-gamma transformation can be determined by more practical means. The maximum likelihood estimate of theta, which is asymptotically unbiased, can be obtained by finding the root in theta of the following likelihood equation given by Birnbaum (1968, p. 459):

$$\sum_{g=1}^{n} a_g \Psi [Da_g(\hat{\theta}-b_g)] - \sum_{g=1}^{n} \frac{w_g(\hat{\theta})\, u_g}{D} = 0 \qquad [3]$$

where

$$w_g(\hat{\theta}) = Da_g \Psi [Da_g(\hat{\theta}-b_g) - \ln(c_g)] \qquad [4]$$

are *locally best* weights as defined by Birnbaum (1968, pp. 442-444), $\hat{\theta}$ is an estimate of θ, and all other terms are as defined earlier, except there is no subscript indexing each examinee.

If each item were repeated r times, Equation 3 could be written as:

$$\sum_{g=1}^{n} \sum_{h=1}^{r} a_g \Psi[Da_g(\hat{\theta}-b_g)] - \sum_{g=1}^{n} \sum_{h=1}^{r} \frac{w_g(\hat{\theta})u_{gh}}{D} = 0 \qquad [5]$$

or

$$r \sum_{g=1}^{n} a_g \Psi [Da_g(\hat{\theta}-b_g)] - \sum_{g=1}^{n} \frac{w_g(\hat{\theta})}{D} \sum_{h=1}^{r} u_{gh} = 0 \qquad [6]$$

or

$$\sum_{g=1}^{n} a_g \Psi[Da_g(\hat{\theta}-b_g)] - \sum_{g=1}^{n} \frac{w_g(\hat{\theta})}{D} P_g = 0 \qquad [7]$$

where P_g = the observed proportion of correct responses to item g in r repetitions.

This is simply Equation 3 with P_g substituted for U_g. If the three-parameter model holds and true parameters are available, P_g can be computed using Equation 1. When this P_g is substituted into Equation 7 along with true item parameters, the root of the equation is found at $\hat{\theta} = \theta$.

In a simulation, we can control when the true parameters are available and when estimates must be used. Let us view P_g as the probability with which an examinee will respond to item g with a correct answer. The probability of an examinee's response being correct is governed by his or her true ability and the true item parameters. Thus, P_g should be computed using θ and a_g, b_g, and c_g. When we estimate ability (i.e., in a real-world environment), we must use the estimated parameters. If the parameters are in error, our estimate ($\hat{\theta}$) will not converge on θ but rather on Γ. The value of Γ corresponding to a given θ can be determined by substituting the true P_g into Equation 7 and finding the root using the errant parameters. Thus, using θ to denote the true ability; Γ to denote the asymptotic value obtainable with errant parameters; a_g, b_g, and c_g to denote the true parameters; and $\hat{a}_g$, $\hat{b}_g$, and $\hat{c}_g$ to denote the estimated parameters, we can rewrite Equation 7 as:

$$\sum_{g=1}^{n} \hat{a}_g \Psi[D\hat{a}_g(\Gamma - \hat{b}_g)] - \sum_{g=1}^{n} \frac{\hat{w}_g(\Gamma)}{D} * \qquad [8]$$

$$(c_g + (1-c_g) \Psi[Da_g(\theta-b_g)]) = 0$$

where

$$\hat{w}_g(\Gamma) = D\hat{a}_g \Psi[D\hat{a}_g(\Gamma - \hat{b}_g) - ln(\hat{c}_g)] \qquad [9]$$

If the errors of calibration are zero or the estimated parameters are consistent with the true parameters, the transformation of theta to gamma will be linear. When this is not the case, as in almost all real calibration situations, the transformation will be nonlinear. This transformation from theta to gamma completely describes the asymptotic effect of item parameter error on ability estimation.

Efficiency. The information at theta for a specific test score (or scoring function), X, can be expressed as the ratio of (1) the squared derivative of the expected value of the scoring function, to (2) the variance of the scoring function at theta (Birnbaum, 1968, p. 453):

$$I(\theta) = I(\theta;\ X) = \frac{\left[\frac{d}{d\theta}\ E(X|\theta)\right]^2}{\sigma^2_{X|\theta}} \qquad [10]$$

When the score is a linear combination of 0-1 item responses, the components of the information equation can be written as:

$$\frac{d}{d\theta}\ E(X|\theta) = \sum_{g=1}^{n} \frac{d}{d\theta}\ w_g\ E(U_g|\theta) \qquad [11]$$

$$= \sum_{g=1}^{n} \frac{d}{d\theta}\ w_g\ P_g(\theta)$$

$$= \sum_{g=1}^{n} w_g\ P'_g(\theta)$$

and

$$\sigma^2_{X|\theta} = \sum_{g=1}^{n} w_g^2\ P_g(\theta)Q_g(\theta) \qquad [12]$$

where $w_g = w_g(\theta)$ is defined as in Equation 4; and

$$P'_g(\theta) = (1-c_g)\ Da_g\ \Psi[Da_g(\theta-b_g)]. \qquad [13]$$

The information available from scoring response vectors using errant parameters can be viewed as equivalent to the information available in a linear combination of item responses using those weights determined to be locally best at Γ, using the estimates of the item parameters. Thus, substituting Equations 1 and 13 and the errant weights from Equation 9 into Equation 10, the information available from the errant parameters is given by Equation 14.

$$I(\theta;\ \hat{a},\ \hat{b},\ \hat{c}) = \frac{\sum_{g=1}^{n} \left[\hat{w}_g(\Gamma)\ P'_g(\theta)\right]^2}{\sum_{g=1}^{n} \hat{w}_g^2\ (\Gamma)\ P_g(\theta)\ Q_g(\theta)}. \qquad [14]$$

Equation 14 represents the information contained in the errant parameters as a function of θ. To produce a single-quantity estimate of the information available, this function may be jointly integrated with a standard normal density function (i.e., numerically integrated).

To provide a relative efficiency index, the information thus obtained may be compared to information available from the parameters. This information may be computed in the same manner using true parameters throughout, or it may be computed using any of the formulas provided by Birnbaum (1968).

Efficiencies for this study were computed in the manner described above. Efficiency, as reported herein, refers to the ratio of errant to true information.

Results

RMSE. Table 2 presents the root mean squared error between the true and estimated parameters for Tests 1, 2, and 3. For each of the tests and each of the calibration programs, the *c* parameter had a smaller RMSE than did the other two parameters. The largest RMSEs in each calibration run were observed for the *a* parameter for Tests 1 and 2, and for the *b* parameter for Test 3.

There were only minor differences in parameter estimation error observed between the two versions of ASCAL. The RMSE for estimating the *a* parameter dropped from 0.150 to 0.132 (from Version 2.0 to Version 3.0) for Test 1, and increased from 0.378 to 0.386 for Test 2. In all other cases, the RMSEs were essentially the same for both versions.

On the other hand, there were consistent differences in the RMSEs between ASCAL and LOGIST. In nearly every case, the RMSE for the parameters produced by LOGIST were larger than those produced by either version of ASCAL; for the Test 1 *b* parameters, for example, the RMSE for LOGIST estimates was more than twice as large as the RMSE from ASCAL. The single exception to this occurred for the *a* parameter in Test 3. Here, the LOGIST RMSE was equal to 0.150; ASCAL Versions 2.0 and 3.0 both produced parameters with an RMSE of 0.161.

Table 2
Root Mean Squared Error Between True and Estimated Item Parameters from Calibration Programs ASCAL 2.0, ASCAL 3.0, and LOGIST5 for Tests 1, 2, and 3

	Item Calibration Program		
	ASCAL 2.0	ASCAL 3.0	LOGIST5
Test 1			
a	0.150	0.132	0.231
b	0.088	0.088	0.182
c	0.049	0.049	0.080
Test 2			
a	0.378	0.386	0.497
b	0.190	0.189	0.250
c	0.079	0.079	0.090
Test 3			
a	0.161	0.161	0.150
b	0.176	0.176	0.183
c	0.054	0.055	0.066

Correlations. The product-moment correlations between the true and estimated parameters are presented in Table 3 for each test and each calibration program. These correlations ranged from .596 to .952 for *a*, from .986 to .996 for *b*, and from .377 to .828 for *c*.

Few differences in the correlations were observed between the two versions of ASCAL. The correlations improved for the *a* parameters in Test 1 (from .935 to .952) and for the *c* parameters in Test 2 (from .613 to .620); in all other instances, the results from the two versions of ASCAL were essentially identical.

The correlations between the true and estimated *b* and *c* parameters were lower for LOGIST than for ASCAL across most comparisons of the three tests; these differences were pronounced for the *c* parameters and small for the *b* parameters. However, the differences in the correlations for the *a* parameters were not consistent between the two types of calibration programs. For Test 1, the correlations for the ASCAL-produced *a* parameters were .935 and .952; for LOGIST, this correlation was .848. For Test 3, the *a* parameters from LOGIST correlated more highly with their true values than did the *a* parameters from ASCAL (.934 versus .926 for both versions of ASCAL). There were essentially no differences in the *a* parameter correlations for Test 2.

Table 3
Product-Moment Correlations Between True and Estimated Parameters from Calibration Programs ASCAL 2.0, ASCAL 3.0, and LOGIST5 for Tests 1, 2, and 3

	Item Calibration Program		
	ASCAL 2.0	ASCAL 3.0	LOGIST5
Test 1			
a	0.935	0.952	0.848
b	0.996	0.996	0.986
c	0.826	0.828	0.612
Test 2			
a	0.598	0.598	0.596
b	0.995	0.995	0.991
c	0.613	0.620	0.498
Test 3			
a	0.926	0.926	0.934
b	0.993	0.993	0.990
c	0.463	0.463	0.377

Efficiency. Table 4 presents the efficiency statistics computed for the three calibration programs and the three tests. Calibration efficiency ranged from .971 to .990 in this table and was generally lowest for Test 2 for both LOGIST and ASCAL. There were essentially no differences in calibration efficiency between the two versions of ASCAL; these values were .990 versus .990 for Test 1, .977 versus .975 for Test 2, and .985 versus .984 for Test 3.

Table 4
Efficiency Statistics from Calibration Programs ASCAL 2.0, ASCAL 3.0, and LOGIST5 for Tests 1, 2, and 3

	Item Calibration Program		
	ASCAL 2.0	ASCAL 3.0	LOGIST5
Test 1	0.990	0.990	0.981
Test 2	0.977	0.975	0.971
Test 3	0.985	0.984	0.986

The calibration efficiency for LOGIST was lower than that for ASCAL for Test 1 (.981 versus .990 for both ASCAL versions) and slightly lower for Test 2 (.971 versus .977 and .975). LOGIST's efficiency was slightly higher for Test 3 (.986 versus .985 and .984).

Summary. There were only minor (and inconsequential) differences in the three evaluative criteria between the two versions of ASCAL. The differences between ASCAL and LOGIST were slightly larger than the differences between the two versions of ASCAL; however, even these differences were small.

The only difference between Test 1 and Test 2 in this study was in the range of the *b* parameters (i.e., the Test 1 difficulty parameters were multiplied by 2.0 to produce the difficulty parameters in Test 2). This had a dramatic effect on item calibration for both ASCAL and LOGIST: (1) for all three item parameters, RMSE was substantially larger in Test 2, (2) the correlations between the true and estimated *a* and *c* parameters were substantially smaller in Test 2, and (3) calibration efficiency was lower for Test 2. Only the *b* parameter correlations were unaffected by this manipulation.

DISCUSSION

The results of this evaluation suggest that ASCAL produces parameter estimates that are at least as accurate as those produced by LOGIST. Thus, although different assumptions and different procedures are used in estimating the parameters, the parameters obtained from either program should be considered of equivalent quality.

However, considerations other than accuracy must also be weighed when choosing between LOGIST and ASCAL. Convenience and cost would dictate that ASCAL be chosen because it is very easy to use and runs economically on a personal computer. On the other hand, ASCAL currently has a limit of 100 items and 5,000 examinees; LOGIST's corresponding limits are somewhat larger. Furthermore, since ASCAL is run with virtually no user-specified options, it lacks the flexibility of LOGIST. Finally, the sole purpose of ASCAL is to estimate item parameters. It has no data editing or transformation capability and it does not produce scores or ability estimates for the examinees. If any of these features are required, LOGIST may be a better choice for item calibration. Otherwise, the economy and simplicity of ASCAL appear to make it a better choice.

REFERENCES

Assessment Systems Corporation. (1984). *User's manual for the MicroCAT Testing System* (Research Rep. No. ONR-85-1). St. Paul, MN: Author.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Jensema, C. J. (1976). A simple technique for estimating latent trait mental test parameters. *Educational and Psychological Measurement, 36,* 705-715.

Kolakowski, D., & Bock, R. D. (1973). *LOGOG: Maximum likelihood item analysis and test scoring--Logistic model for multiple item responses (User's Guide).* Ann Arbor, MI: National Educational Resources.

Lord, F. M. (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika, 39,* 247-264.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Ree, M. J. (1978). Estimating item characteristic curves. *Applied Psychological Measurement, 3,* 371-385.

Vale, C. D., Maurelli, V. A., Gialluca, K. A., Weiss, D. J., & Ree, M. J. (1981). *Methods for linking item parameters* (AFHRL-TR-81-10). Brooks Air Force Base, TX: Air Force Human Resources Laboratory.

Vale, C. D., & Weiss, D. J. (1975). *A simulation study of stradaptive ability testing* (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program. (AD A020961)

Wingersky, M. S., Barton, M. A., & Lord, F. M. (1982). *LOGIST user's guide.* Princeton, NJ: Educational Testing Service.

Wood, R. L., Wingersky, M. S., & Lord, F. M. (1976). *LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters* (Research Memorandum 76-6). Princeton, NJ: Educational Testing Service.

Distribution List

Personnel Analysis Division,
AF/MPXA
5C360, The Pentagon
Washington, DC 20330

Air Force Human Resources Lab
AFHRL/MPD
Brooks AFB, TX 78235

Dr. Earl A. Alluisi
HQ, AFHRL (AFSC)
Brooks AFB, TX 78235

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Phipps Arabie
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Technical Director, ARI
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Isaac Bejar
Educational Testing Service
Princeton, NJ 08450

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Tel Aviv, Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaiwes
Code N711
Naval Training Equipment Center
Orlando, FL 32813

Dr. R. Darrell Bock
University of Chicago
Department of Education
Chicago, IL 60637

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code N-095R
NAVTRAEQUIPCEN
Orlando, FL 32813

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. Patricia A. Butler
NIE Mail Stop 1806
1200 19th St., NW
Washington, DC 20208

Mr. James W. Carey
Commandant (G-PTE)
U.S. Coast Guard
2100 Second Street, S.W.
Washington, DC 20593

Dr. James Carlson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd.
Chapel Hill, NC 27514

Dr. Robert Carroll
NAVOP 01B7
Washington, DC 20370

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
University Park
Los Angeles, CA 90007

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans Crombag
University of Leyden
Education Research Center
Boerhaavelaan 2
2334 EN Leyden
The NETHERLANDS

CTB/McGraw-Hill Library
2500 Garden Road
Monterey, CA 93940

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Ball Foundation
800 Roosevelt Road
Building C, Suite 206
Glen Ellyn, IL 60137

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
Attn: TC
(12 Copies)

Dr. Stephen Dunbar
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Kent Eaton
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. John M. Eddins
University of Illinois
252 Engineering Research
Laboratory
103 South Mathews Street
Urbana, IL 61801

Dr. Susan Embretson
University of Kansas
Psychology Department
Lawrence, KS 66045

ERIC Facility-Acquisitions
4833 Rugby Avenue
Bethesda, MD 20014

Dr. Benjamin A. Fairbank
Performance Metrics, Inc.
5825 Callaghan
Suite 225
San Antonio, TX 78228

Dr. Leonard Feldt
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
Program
P.O. Box 168
Iowa City, IA 52240

Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Carl H. Frederiksen
McGill University
3700 McTavish Street
Montreal, Quebec H3A 1Y2
CANADA

Dr. Robert D. Gibbons
University of Illinois-Chicago
P.O. Box 6998
Chicago, IL 69680

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Ronald K. Hambleton
Prof. of Education & Psychology
University of Massachusetts
at Amherst
Hills House
Amherst, MA 01003

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 62
San Diego, CA 92152

Dr. Paul W. Holland
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Prof. Lutz F. Hornke
Universitat Dusseldorf
Erziehungswissenschaftliches
Universitatsstr. 1
Dusseldorf 1
WEST GERMANY

Dr. Paul Horst
677 C Street, #184
Chula Vista, CA 90010

Mr. Dick Hoshaw
NAVOP-135
Arlington Annex
Room 2834
Washington, DC 20350

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
Department of Education
University of Alberta
Edmonton, Alberta
CANADA

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Department of Psychology
University of South Carolina
Columbia, SC 29208

Dr. Douglas H. Jones
Advanced Statistical
Technologies Corporation
10 Trafalgar Court
Lawrenceville, NJ 08148

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
University of Texas-Austin
Measurement and Evaluation
Center
Austin, TX 78703

Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Oude Boteringestraat 23
9712GC Groningen
The NETHERLANDS

Dr. Robert Linn
College of Education
University of Illinois
Urbana, IL 61801

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. James Lumsden
Department of Psychology
University of Western Australia
Nedlands W.A. 6009
AUSTRALIA

Dr. William L. Maloy
Chief of Naval Education
and Training
Naval Air Station
Pensacola, FL 32508

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen Martin
Army Research Institute
5001 Eisenhower Blvd.
Alexandria, VA 22333

Dr. James McBride
Psychological Corporation
c/o Harcourt, Brace,
Javanovich Inc.
1250 West 6th Street
San Diego, CA 92101

Dr. Clarence McCormick
HQ, MEPCOM
MEPCT-P
2500 Green Bay Road
North Chicago, IL 60064

Mr. Robert McKinley
University of Toledo
Department of Educational Psychology
Toledo, OH 43606

Dr. Barbara Means
Human Resources
Research Organization
1100 South Washington
Alexandria, VA 22314

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Headquarters, Marine Corps
Code MPI-20
Washington, DC 20380

Dr. W. Alan Nicewander
University of Oklahoma
Department of Psychology
Oklahoma City, OK 73069

Dr. William E. Nordbrock
FMC-ADCO Box 25
APO, NY 09710

Dr. Melvin R. Novick
356 Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Director, Manpower and Personnel
Laboratory,
NPRDC (Code 06)
San Diego, CA 92152

Distribution List (Continued)

Library, NPRDC
Code P201L
San Diego, CA 92152

Commanding Officer,
Naval Research Laboratory
Code 2627
Washington, DC 20390

Dr. James Olson
WICAT, Inc.
1875 South State Street
Orem, UT 84057

Office of Naval Research,
Code 1142PT
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Special Assistant for Marine
Corps Matters,
ONR Code 00MC
800 N. Quincy St.
Arlington, VA 22217-5000

Dr. Judith Orasanu
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dr. Roger Pennell
Air Force Human Resources
Laboratory
Lowry AFB, CO 80230

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MP
Brooks AFB, TX 78235

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
Knoxville, TN 37916

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152

Dr. Robert Sasmor
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Mary Schratz
Navy Personnel R&D Center
San Diego, CA 92152

Dr. W. Steve Sellman
OASD(MRA&L)
2B269 The Pentagon
Washington, DC 20301

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujusawa 251
JAPAN

Dr. William Sims
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. H. Wallace Sinaiko
Manpower Research
and Advisory Services
Smithsonian Institution
801 North Pitt Street
Alexandria, VA 22314

Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Paul Speckman
University of Missouri
Department of Statistics
Columbia, MO 65201

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
200 North Beauregard Street
Alexandria, VA 22311

Dr. William Stout
University of Illinois
Department of Mathematics
Urbana, IL 61801

Maj. Bill Strickland
AF/MPXOA
4E168 Pentagon
Washington, DC 20330

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Kikumi Tatsuoka
CERL
252 Engineering Research
Laboratory
Urbana, IL 61801

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
The Fred Hutchinson
Cancer Research Center
Division of Public Health Sci.
1124 Columbia Street
Seattle, WA 98104

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. Vern W. Urry
Personnel R&D Center
Office of Personnel Management
1900 E. Street, NW
Washington, DC 20415

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 310
St. Paul, MN 55114

Dr. Frank Vicino
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Howard Wainer
Division of Psychological Studies
Educational Testing Service
Princeton, NJ 08541

Dr. Ming-Mei Wang
Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Mr. Thomas A. Warm
Coast Guard Institute
P. O. Substation 18
Oklahoma City, OK 73169

Dr. Brian Waters
Program Manager
Manpower Analysis Program
HumRRO
1100 S. Washington St.
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455

Dr. Ronald A. Weitzman
NPS, Code 54Wz
Monterey, CA 92152

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Rand R. Wilcox
University of Southern
California
Department of Psychology
Los Angeles, CA 90007

German Military Representative
ATTN: Wolfgang Wildegrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Army Research Institute
5001 Eisenhower Ave.
Alexandria, VA 22333

Dr. Martin F. Wiskoff
Navy Personnel R & D Center
San Diego, CA 92152

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering
Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940